

Wavelet smoothing of fMRI activation images

Will Penny,
Wellcome Department of Imaging Neuroscience,
University College, London WC1N 3BG.
wpenny@fil.ion.ucl.ac.uk

October 25, 2002

1 Introduction

Images from functional imaging experiments are subject to a data processing stream involving motion correction and spatial normalization. The next step is then to smooth the images using a Gaussian kernel. One reason for the smoothing step is to render activation information amenable to classical statistical inference via Random Field Theory. This requires that the residual fields have a smoothness, measured by the Full Width at Half Maximum (FWHM) of a Gaussian kernel, that is of the order of three times the voxel edge size.

If one were to abandon the classical inference framework in favour of Bayesian inference however, see eg. [5], then the above motivation for smoothing disappears. This then leaves the second reason for smoothing which derives from the fact that a point mass of neural activity gives rise to a spatially extended hemodynamic response, the relation being defined by the spatial hemodynamic Point Response Function (PRF). If one assumes that the PRF is Gaussian with a known FWHM, then one can increase the Signal-to-Noise Ratio (SNR) in functional images by smoothing with that kernel. Over- or under specification of the FWHM, however, will lead to suboptimal increase in SNR. Further, if the PRF is non-stationary then smoothing with a fixed size Gaussian kernel is clearly sub-optimal. This provides the motivation for smoothing with other methods - such as wavelets.

2 Wavelet Transform

The wavelet spatial model is that introduced by Mallat [8] and described further in [9]. It consists of a multiresolution hierarchy in which an image is represented at a number of spatial resolutions. These are known as the 'coarse' levels where lower levels correspond to successively lower frequency aspects of the original image. The difference between successive coarse level images are the 'detail' images. These correspond to high frequency components. Overall, a wavelet transform is defined in which an N -dimensional image is represented by an N -dimensional image of wavelet coefficients. These coefficients constitute the coarse and detail levels making up the multiresolution hierarchy, as shown in standard figures (see eg. figures 10 and 11 in [8]).

Importantly, the wavelet transform (WT) is orthogonal and there exists a fast 'cascade' algorithm for implementing it. An image can be exactly reconstructed using a fast inverse wavelet transform (IWT). Wavelets are of interest to us because it turns out that, for images of natural scenes, the probability density over the wavelet coefficients is heavy-tailed (super-Gaussian) ie. the vast majority of coefficients are close to zero and a few are large. For Gaussian white noise images, however, the probability density is itself Gaussian. Hence, natural images which have been corrupted by additive Gaussian white noise can be restored by forcing the wavelet coefficient probability density to be heavy tailed and then applying the IWT. This can be achieved with Bayesian inference using shrinkage priors.

We consider images having $N_x \times N_y$ pixels. These are then represented as N -element column vectors v where $N = N_x N_y$. The corresponding vector of wavelet coefficients d (which can be re-shaped into an image) can, in principle, be computed by projecting v onto a set of wavelet basis functions, W , ie.

$$d = W^T v \tag{1}$$

In practice, this projection takes place using a fast cascade algorithm (see [11] for an excellent description). Equivalently, the original image is represented using

$$v = W d \tag{2}$$

where we make use of the orthogonality of the WT ie. $W^T W = I$. At each level in Mallat's multiresolution hierarchy there are coefficients referring to three different orientations ('horizontal', 'vertical' and 'horizontal and verti-

cal'). So, for a 4-resolution hierarchy there are 12 sub-images in the wavelet coefficient images. If we let j index these sub-images and k index the k th wavelet coefficient within each sub-image we can write

$$v = \sum_{j,k} W(j,k)d_{jk} \quad (3)$$

where $W(j,k)$ is a wavelet basis function picked out from the appropriate column of W and d_{jk} is a scalar.

3 Shrinkage Priors

In this paper we consider the shrinkage prior described by Clyde and George [3] (see also [2]). This assumes that the wavelet coefficients are a priori independent and the prior density of each coefficient is given by a mixture of a zero-mean Gaussian and a point mass at zero. The parameters of the prior distributions are tuned for each detail level j . The prior probability of a coefficient being non-zero is α_j and the variance of the Gaussian is $c_j\sigma$ where σ is the standard deviation of the observation noise (this necessarily assumes that the observation noise variance is uniform). We note that shrinkage is only applied to the detail levels and not the coarse level.

Due to the orthogonality of the basis functions, the fact that the additive noise is white, and the independence of the prior, the posterior distribution of wavelet coefficients also factorises over coefficients. The posterior probability that the wavelet coefficient is non-zero is γ_{jk} . This probability and the parameters α_j and c_j can be computed using the following Empirical Bayes algorithm [3].

$$\begin{aligned} p_{jk} &= \frac{1}{\sqrt{1+c_j}} \frac{\alpha_j}{1-\alpha_j} \exp\left(0.5 \left(\frac{d_{jk}}{\sigma}\right)^2 \frac{c_j}{1+c_j}\right) \\ \gamma_{jk} &= \frac{p_{jk}}{1+p_{jk}} \\ c_j &= \frac{\sum_k \gamma_{jk} d_{jk}^2}{\sigma^2 \sum_k \gamma_{jk}} - 1 \\ \alpha_j &= \frac{\sum_k \gamma_{jk}}{N_j} \end{aligned} \quad (4)$$

where N_j are the number of wavelet coefficients in the j th sub-image. We start with the values $\alpha_j = 0.1, c_j = 0$ and the equations are iterated until consistent values for α_j and c_j are reached.

The posterior mean of the j, k th coefficient is then given by

$$s_{jk} = \frac{c_j}{1 + c_j} \gamma_{jk} d_{jk} \quad (5)$$

Two different point estimates of the wavelet coefficients are then made. The first is the ‘threshold’ estimate

$$t_{jk} = \frac{c_j}{1 + c_j} (\gamma_{jk} > 0.5) d_{jk} \quad (6)$$

which sets the value either to zero or to $d_{jk} c_j / (1 + c_j)$. The second is the ‘model averaging’ point estimate

$$a_{jk} = s_{jk} \quad (7)$$

so called because it averages over the posterior probability of mixture components. The IWT is then applied to either t_{jk} or a_{jk} to restore the image. We refer to the above restoration processes as the Bayesian thresholding and Bayesian averaging methods. As we shall see, the averaging methods are more accurate but the thresholding methods require fewer basis functions.

4 Results

The following applications used the bi-orthogonal wavelet bases described by Daubechies [9].

4.1 Dexter image

Figure 1 shows a 4-level multiresolution representation of a natural image (containing Professor Denzil Dexter, University of Southern California). The wavelet coefficients correspond to the lowest coarse level and the detail images (not shown). We then added Gaussian noise to the original image and applied the Bayesian restoration algorithms. As expected the Bayesian averaging method provides the best reconstruction in terms of squared error. The reconstruction error of the Bayesian threshold methods was marginally bigger but this was achieved with less than a full basis set. In fact, at detail levels 4 to 7 only 60%, 25%, 5% and 1% of coefficients were kept (these are typical values across the three different orientations at each level).

4.2 fMRI activation images

We now show how Bayesian wavelet restoration can be used to smooth contrast images from functional imaging experiments. We are therefore considering smoothing at a later stage of the data processing stream than is usual. Unsmoothed but motion-corrected and spatially normalised functional images are modelled using a voxel-wise General Linear Model (GLM) approach. In this paper we used the algorithm described in [10]. One then specifies a contrast vector to test for the effect of scientific interest. In this paper we have analysed data from a face processing study [6] and chosen a contrast vector that looks for voxels that are differentially active when presented with face images versus checkerboard images. We look at a single transverse slice of data from 12 different subjects. Application of the contrast vector to the GLM produces an image of the estimated effect size (the so-called 'con' image), v_{xy} . This is accompanied by an estimate of the effect variability ie. a variance image, σ_{xy}^2 . The two are related as follows

$$v_{xy} = h_{xy} + e_{xy} \quad (8)$$

where $\text{Var}[e_{xy}] = \sigma_{xy}^2$ and h_{xy} is the true effect size. That is, for voxels with larger σ_{xy}^2 we are more uncertain about the true effect size (Note that as we've explained it, equation 8 should be $h_{xy} = v_{xy} + e_{xy}$ but the two equations are the same except for the sign of the error). The key point is that, mathematically, the problem of estimation of the true effect size is identical to the problem of restoration of images corrupted by additive noise. We can therefore use image restoration algorithms to solve our effect size problem and, fortunately, this is a mature field with a large body of literature. Wavelet methods are our current focus.

We first note that the standard algorithms are not readily applied to images with heteroscedastic noise ie. nonstationary error variance. This is because the posterior distribution over wavelet coefficients no longer factorises. One can, however, resort to methods described in [7] in which bandlimited covariance matrices are computed using wavelet transforms. For images, however, these matrices are still prohibitively large.

An alternative solution is to apply the wavelet transform to contrast images which have been normalised with respect to the error variance at each voxel. These pre-processed data then have constant error variance σ_g ie.

$$\tilde{v}_{ij} = \sigma_g \frac{v_{ij}}{\sigma_{ij}} \quad (9)$$

We then apply the wavelet restoration process on \tilde{v}_{ij} to get restored images \tilde{r}_{ij} . This amounts to smoothing normalised effect-size images (or, normalised contrast images).

Normalised contrast images from each of 12 subjects are shown in Figure 3. These have been produced by a voxel-wise GLM analysis. Figure 4 shows the result of smoothing these images using a fixed width Gaussian kernel and Figure 5 shows the wavelet restored images using the Bayesian averaging approach. All images in these figures are on the same scale. The signal size in activation areas is greater for the wavelet-smoothed images. Gaussian smoothing blurs these signals resulting in a lower SNR. The main reason for this is that the Gaussian approach uses stationary smoothing whereas the wavelet approach uses nonstationary smoothing (ie. effectively with a kernel of varying width).

5 Discussion

Spatial wavelets have been used in the analysis of PET images by Turkheimer et al. [12]. This work used Battle-Lemarie wavelets and a wavelet thresholding based on minimax [4] and Bonferroni criterion. Brammer [1] applied spatio-temporal wavelets to fMRI images using the bi-orthogonal Daubechies bases. Significantly non-zero clusters of wavelet coefficients were identified using Kolmogorov-Smirnoff statistics.

Both of the above approaches estimate the noise variance by looking at the variability of wavelet coefficients at the finest level of detail (this is the standard procedure in wavelet restoration). In our approach we used an estimate based on a global value, σ_g^2 , derived from the average noise variance across subjects. This quantifies the average uncertainty of the normalised effect-size at each voxel.

Following the logic of our approach to random effects analysis (ie. looking at the mean and variance images across subjects) yielded activation images with very low spatial resolution. One inconsistency in our approach is that we have smooth signal images but unsmooth variance images. A fully consistent approach would have a spatial model for the signal and the error. Finally, we would ideally like to look at the posterior distribution of effect size in image space. Whilst this can be achieved with wavelets, it is computationally prohibitive (see references in section 3.2 of [2]).

References

- [1] M.J. Brammer. Multidimensional wavelet analysis of functional magnetic resonance images. *Human Brain Mapping*, 6:378–382, 1998.
- [2] M. Clyde, G. Parmigiani, and B. Vidakovic. Multiple shrinkage and subset selection selection in wavelets. *Biometrika*, 85:391–402, 1998.
- [3] M.A. Clyde and E. I. George. Empirical Bayes estimation in wavelet nonparametric regression. Technical Report 99-06, Institute of Statistics and Decision Sciences. Working paper series., 1999. Available from <http://www.stat.duke.edu/papers/working-papers-99.html>.
- [4] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.
- [5] K.J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner. Classical and Bayesian inference in neuroimaging: Theory. *Neuroimage*, 16:465–483, 2002.
- [6] R.N.A. Henson, T. Shallice, M.L. Gorno-Tempini, and R.J. Dolan. Face repetition effects in implicit and explicit memory tests as measured by fMRI. *Cerebral Cortex*, 12:178–186, 2002.
- [7] A. Kovac and B.W. Silverman. Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *Journal of the American Statistical Association*, 95(449):172–183, March 2000.
- [8] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11:674–693, 1989.
- [9] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1999.
- [10] W.D. Penny, S.J. Kiebel, and K.J. Friston. Variational bayesian inference for fmri time series. Technical report, Submitted to Neuroimage, 2002.
- [11] W. H. Press, S.A. Teukolsky, W.T. Vetterling, and B.V.P. Flannery. *Numerical Recipes in C*. Cambridge, 1992.

- [12] F.E. Turkheimer, M. Brett, J.A.D. Aston, A.P. Leff, P.A. Sargent, R.J. Wise, P.M. Grasby, and V.J. Cunningham. Statistical modelling of positron emission tomography images in wavelet space. *Journal of Cerebral Blood Flow and Metabolism*, 20:1610–1618, 2000.

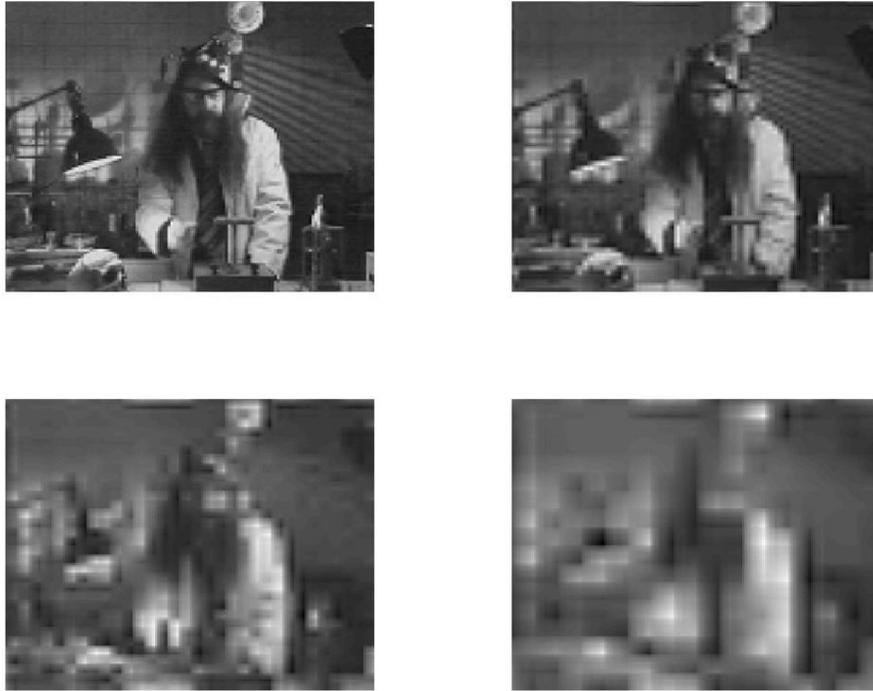


Figure 1: *Coarse images at various resolution levels, L , where (a) $L = 7$, (b) $L = 6$, (c) $L = 5$ and (d) $L = 4$. At resolution level L the image is represented by 2^L pixels in each direction. In the above figure the images been reshaped to the same size. The detail images at a particular level are produced by 'horizontal', 'vertical' and 'horizontal and vertical' differences between successive levels. The set of coefficients produced by the wavelet transform consist of the lowest coarse level images and the higher level detail images. When added together these reproduce the original image without error. The original image is $2^8 \times 2^8$ pixels and is shown in figure 2(a).*

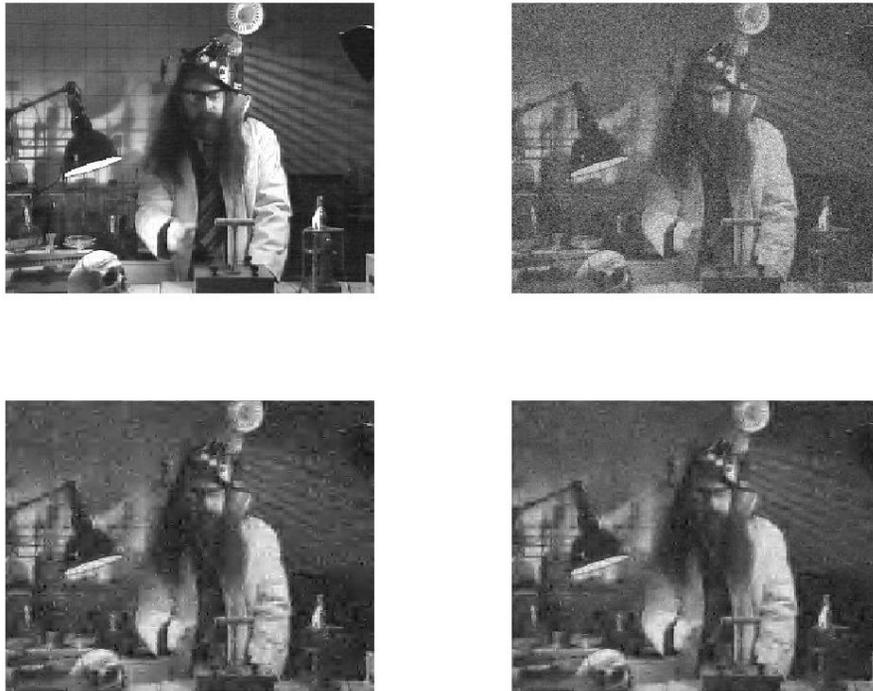


Figure 2: (a) Original image, (b) corrupted with additive Gaussian noise of known variance (Mean Squared Error=0.36), (c) wavelet-restored image using Bayesian thresholding (Error=0.086), (d) wavelet-restored image using Bayesian averaging (Error=0.072).

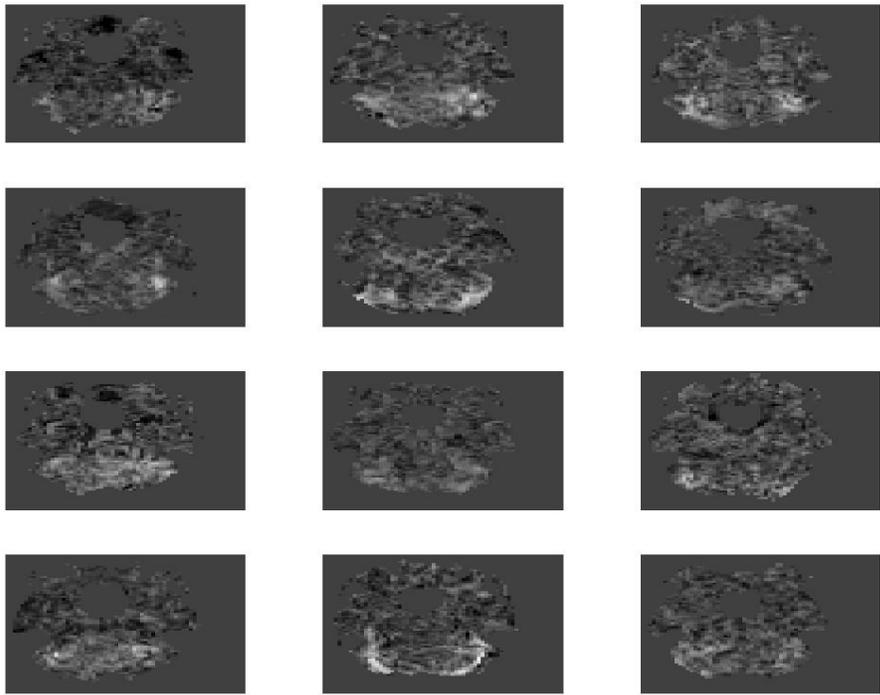


Figure 3: *Contrast images from 12 subjects.*

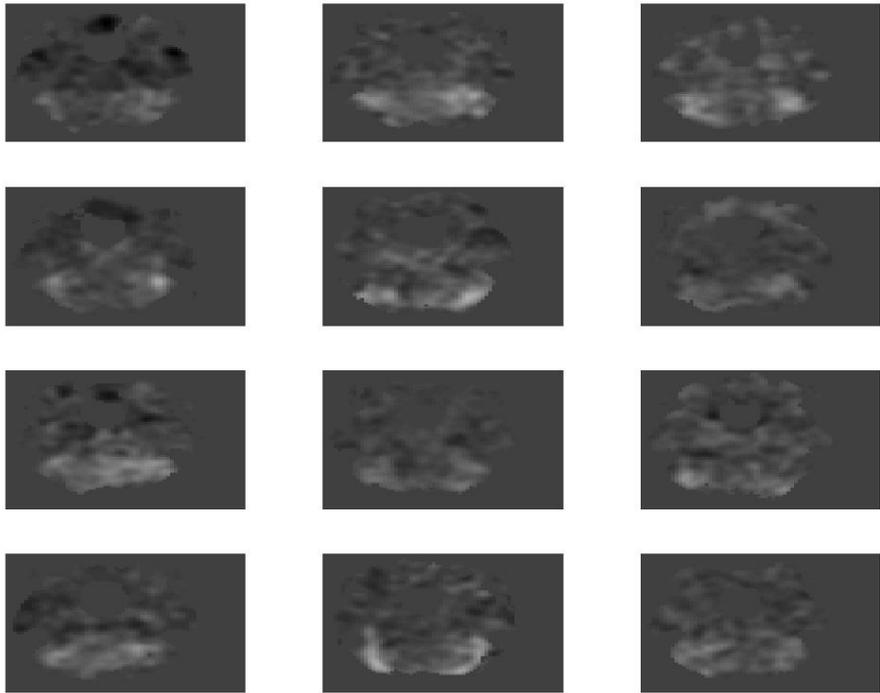


Figure 4: *Gaussian smoothed contrast images from 12 subjects (FWHM=8mm).*

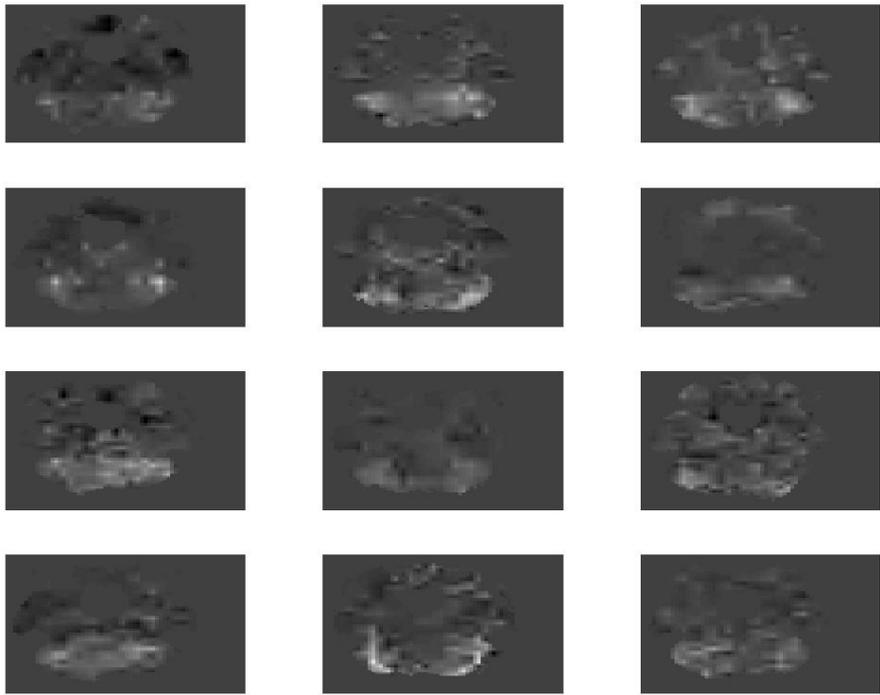


Figure 5: *Wavelet restored contrast images from 12 subjects.*